

# BIRCH, STEWART, KOLASCH & BIRCH, LLP

TERRELL C BIRCH  
RAYMOND C STEWART  
JOSEPH A KOLASCH  
JAMES M SLATTERY  
BERNARD L SWEENEY\*  
MICHAEL K MUTTER  
CHARLES GORENSTEIN  
GERALD M MURPHY, JR  
LEONARD R SVENSSON  
TERRY L CLARK  
ANDREW D MEIKLE  
MARC S WEINER  
JOE MCKINNEY MUNCY  
ROBERT J KENNEY  
DONALD J DALEY  
JOHN W BAILEY  
JOHN A CASTELLANO, III

OF COUNSEL:  
HERBERT M BIRCH (1905-1996)  
ELLIOT A GOLDBERG\*  
WILLIAM L GATES\*  
EDWARD H VALANCE  
RUPERT J BRADY (RET.)\*  
\*ADMITTED TO A BAR OTHER THAN VA

INTELLECTUAL PROPERTY LAW  
8110 GATEHOUSE ROAD  
SUITE 500 EAST  
FALLS CHURCH, VA 22042-1210  
USA

(703) 205-8000

FAX: (703) 205-8050  
(703) 698-8590 (G IV)

e-mail: mailroom@bskb.com  
web: http://www.bskb.com

CALIFORNIA OFFICE  
650 TOWN CENTER DRIVE, SUITE 1120  
COSTA MESA, CA 92626-7125

GARY D YACURA  
THOMAS S AUCHTERLONIE  
MICHAEL R CAMMARATA  
JAMES T ELLER, JR  
SCOTT L LOWE  
MARY ANN CAPRIA  
MARK J NUELLE, PH.D  
DARIN E BARTHOLOMEW  
D RICHARD ANDERSON  
PAUL C LEWIS  
W KARL RENNER  
MARK W MILSTEAD\*  
JOHN CAMPA\*

REG PATENT AGENTS:  
FREDERICK R. HANDRES  
ANDREW J TELESZ, JR  
MARYANNE ARMSTRONG, PH.D.  
MIKE HATSUMI  
MIKE S RYU  
CRAIG A. McROBBIE  
GARTH M DAHLEN, PH.D.  
LAURA C LUTZ  
ROBERT E. GOOZNER, PH.D.  
HYUNG N SOHN  
MATTHEW J. LATTIG  
ALAN PEDERSEN-GILES  
JUSTIN D KARJALA

Date: February 25, 2000

Docket No.: 2750-198P

## BOX PATENT APPLICATION

Assistant Commissioner for Patents  
Washington, DC 20231

Sir:

As authorized by the inventor(s), transmitted herewith for filing is a patent application applied for on behalf of the inventor(s) according to the provisions of 37 C.F.R. § 1.41(c), which claims priority under 35 U.S.C. § 119(e) of Provisional Application No. 60/121,700 filed on February 25, 1999.

Inventor(s): K. Diane JOFUKU and Anne-Marie BOUCKAERT

For: METHODS OF ISOLATING AND/OR IDENTIFYING RELATED PLANT SEQUENCES

Enclosed are:

- ☒ A specification consisting of forty-one (41) pages
- ☐ ( ) sheet(s) of formal drawings
- ☐ Certified copy of Priority Document(s)
- ☒ Executed Declaration in accordance with 37 C.F.R. § 1.64 will follow
- ☒ A statement to establish small entity status under 37 C.F.R. § 1.9 and 37 C.F.R. § 1.27
- ☐ Preliminary Amendment

- ☒ Information Sheet
- ☐ Information Disclosure Statement, PTO-1449 and reference(s)
- ☒ Amend the specification by inserting before the first line the sentence:

--This application claims priority on provisional Application No. 60/121,700 filed on February 25, 1999, the entire contents of which are hereby incorporated by reference.--

- ☒ Other: Power of Attorney for Claiming Small Entity Status

The filing fee has been calculated as shown below:

			LARGE ENTITY	SMALL ENTITY
BASIC FEE			\$690.00	\$345.00
	NUMBER FILED	NUMBER EXTRA	RATE FEE	RATE FEE
<b>TOTAL CLAIMS</b>	46- 20 =	26	X 18 = \$0.00	x 9 = \$234.00
<b>INDEPENDENT CLAIMS</b>	12- 3 =	9	x 78 = \$0.00	x 39 = \$351.00
<input checked="" type="checkbox"/> MULTIPLE DEPENDENT CLAIMS PRESENTED			+ \$260.00	+ \$130.00
<b>TOTAL</b>			\$0.00	\$1,060.00

- ☒ The application transmitted herewith is filed in accordance with 37 C.F.R. § 1.41(c). The undersigned has been authorized by the inventor(s) to file the present application. The original duly executed declaration together with the surcharge will be forwarded in due course.
- ☒ A check in the amount of \$1,060.00 to cover the filing fee is enclosed.
- ☐ Please charge Deposit Account No. 02-2448 in the amount of \$0.00. A triplicate copy of this transmittal form is enclosed.

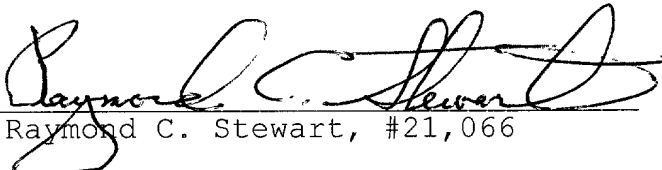
☒ Please send correspondence to:

BIRCH, STEWART, KOLASCH & BIRCH, LLP **or** Customer No. 2292  
P.O. Box 747  
Falls Church, VA 22040-0747  
Telephone: (703) 205-8000

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. §§ 1.16 or 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By   
Raymond C. Stewart, #21,066

P.O. Box 747  
Falls Church, VA 22040-0747  
(703) 205-8000

RCS/DRN/las  
2750-198P

Attachments

(Rev. 01/08/2000)

STATEMENT CLAIMING SMALL ENTITY STATUS  
(37 CFR 1.9(f) & 1.27(c)) - SMALL BUSINESS CONCERN

Docket Number: 2750-198P

Applicant, Patentee, or Identifier: K. Diane JOFUKU and Anne-Marie BOUCKAERT

Application or Patent No.: NEW Provisional

Filed or Issued: February 25, 2000

Title: METHODS OF ISOLATING AND/OR IDENTIFYING RELATED PLANT SEQUENCES

I hereby state that I am

- ☐ the owner of the small business concern identified below:  
☒ an official of the small business concern empowered to act on behalf of the concern identified below:

NAME OF SMALL BUSINESS CONCERN CERES, INC.

ADDRESS OF SMALL BUSINESS CONCERN  
3007 Malibu Canyon Road Malibu, CA 90265

I hereby state that the above identified small business concern qualifies as a small business concern as defined in 37 CFR Part 121 for purposes of paying reduced fees to the United States Patent and Trademark Office, in that the number of employees of the concern, including those of its affiliates, does not exceed 500 persons. For purposes of this statement, (1) the number of employees of the business concern is the average over the previous fiscal year of the concern of the persons employed on a full-time, part-time, or temporary basis during each of the pay periods of the fiscal year, and (2) concerns are affiliates of each other when either, directly or indirectly, one concern controls or has the power to control the other, or a third party or parties controls or has the power to control both.

I hereby state that rights under contract or law have been conveyed to and remain with the small business concern identified above with regard to the invention described in:

- ☒ the specification filed herewith with title as listed above.  
☐ the application identified above.  
☐ the patent identified above.

If the rights held by the above identified small business concern are not exclusive, each individual, concern, or organization having rights in the invention must file separate statements as to their status as small entities, and no rights to the invention are held by any person, other than the inventor, who would not qualify as an independent inventor under 37 CFR 1.9(c) if that person made the invention, or by any concern which would not qualify as a small business concern under 37 CFR 1.9(d), or a nonprofit organization under 37 CFR 1.9(e).

Each person, concern, or organization having any rights in the invention is listed below:

- ☒ no such person, concern, or organization exists.  
☐ each such person, concern, or organization is listed below.

Separate statements are required from each named person, concern, or organization having rights to the invention stating their status as small entities. (37 CFR 1.27)

I acknowledge the duty to file, in this application or patent, notification of any change in status resulting in loss of entitlement to small entity status prior to paying, or at the time of paying, the earliest of the issue fee or any maintenance fee due after the date on which status as a small entity is not longer appropriate. (37 CFR 1.28(b))

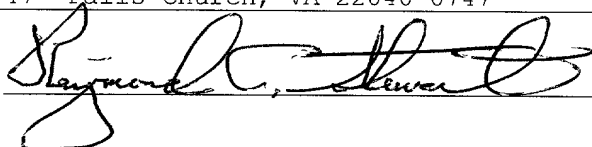
NAME OF PERSON SIGNING Raymond C. Stewart (Reg. No. 21,066)

TITLE IN ORGANIZATION OF PERSON SIGNING Legal Representative of CERES, INC.

ADDRESS OF PERSON SIGNING Birch, Stewart, Kolasch and Birch, LLP.

P.O. Box 747 Falls Church, VA 22040-0747

SIGNATURE



DATE February 25, 2000

**METHODS OF ISOLATING AND/OR IDENTIFYING RELATED  
PLANT SEQUENCES**

**FIELD OF THE INVENTION**

This invention is related to utilizing molecular biology and recombinant DNA technology to isolate and/or identify sequences from different plant families.

5 **BACKGROUND OF THE INVENTION**

References describing codon usage include: Carels et al., *J. Mol. Evol.*, Vol. 46, pp. 45-53 (1998) and Fennoy et al., *Nucl. Acids Res.*, Vol. 21, No. 23, pp. 5294-5300 (1993).

10 AP2 like proteins and genes of *Arabidopsis* re described in copending U.S. Application Nos. 08/700,152; 08/879,827; 08/912,272; and 09/026,039.

**SUMMARY OF THE INVENTION**

15 The present invention relates to a method of isolating a target polynucleotide from a target plant species that encodes a polypeptide exhibiting a desired degree of sequence identity to a conserved region of a template polypeptide from a template plant species. The method comprises:

20 (a) identifying the amino acid sequence of the conserved region in the template polypeptide;

(b) generating an oligonucleotide comprising a sequence wherein the sequence or its reverse complement encodes at least four amino acids of the conserved region  
25 identified in step (a), wherein

(i) the nucleotide of the first and second position of at least three codons are the same as the

corresponding nucleotides in the template polynucleotide encoding the template polypeptide; and

(ii) the nucleotide of the third position of the codon of step (i) is the same as the nucleotide at the third position of the most preferred codon of the second plant class, family, genera, or species for that amino acid in the portion of the conserved region;

further wherein the oligonucleotide preferably does not comprise homopolymers of more than four nucleotides; and the oligonucleotide is not degenerate;

(c) providing a composition comprising the target polynucleotide;

(d) contacting the oligonucleotide and the target polynucleotide under conditions that permit hybridization and formation of a duplex.

Identification of target polynucleotide can be accomplished by detection of the duplex of step (d). Further, both single stranded and double stranded target polynucleotides can be generated from the duplex of step (d).

## **DETAILED DESCRIPTION OF THE INVENTION**

### **Definitions**

The usage of the term "plant family" herein refers to the common nomenclature used to classify organisms, for example Liliaceae and Orchidaceae are plant families.

### **General Method**

The present invention relates to a method of isolating and/or identifying genes in nucleic acids from a target plant species related to a gene or corresponding cDNA or other nucleic acids from a template plant species.

Preferably, the target and template plant species are from different plant families.

In another embodiment of the invention, the method includes identifying and/or isolating from a target plant  
5 species a target polynucleotide that encodes a conserved region that exhibits at least 70% sequence a conserved region encoded by the template polynucleotide from another plant species.

The target and template polynucleotides can be either  
10 RNA or DNA or derivatives thereof. The oligonucleotides to be utilized can be RNA, DNA, or derivatives thereof, such as protein-nucleic acids, (PNAs). The target polynucleotide can be isolated from cDNA or genomic libraries or fixed on microarrays and need not be isolated directly from the  
15 second or target plant organism. Such plant sequences can be first subcloned into intermediary vectors or organisms.

The method utilizes sequences from a conserved region of the polypeptide encoded by the template polynucleotide. A "conserved region" is a primary sequence within a  
20 polypeptide that correlates to an *in vitro* activity, *in vivo* activity, or a secondary structure. For example, the active site of a serine protease exhibits a particular tertiary structure that is responsible for the activity of the protein. That same tertiary structure can be encoded by way  
25 of different amino acid sequences, but certain portions of the sequence tend to be the same among the variants. The amino acid sequence identity of conserved regions from related proteins can be as low as approximately 35%. Thus, even polypeptides that exhibit about 35% sequence identity  
30 can be useful to identify a conserved region. More typically, such conserved regions of related proteins exhibit at least 50% sequence identity; even more typically

at least about 60%; even more typically, at least 70% sequence identity, more typically at least 80%, even more typically about 90% sequence identity.

A. Identifying Conserved Regions

5 Conserved regions can be identified by locating a primary sequence within the template polypeptide that:

- (i) is a repeated sequence;
- (ii) forms some secondary structure, such as helices, beta sheets, etc.
- 10 (iii) establishes positively or negatively charged domains;

(iv) represent a protein motif or domain. See, for example, the Pfam web site describing the consensus sequence for a variety of protein motifs and domains. The sites on  
15 the World Wide Web in the UK at <http://www.sanger.ac.uk/Pfam/> and in the US at <http://genome.wustl.edu/Pfam/>. For a description of the information included at the Pfam database, see Sonnhammer et al., Nucl Acids Res 26(1): 320-322 (January 1, 1998); and  
20 Sonnhammer EL, Eddy SR, Durbin R (1997) Pfam: A Comprehensive Database of Protein Families Based on Seed Alignments, Proteins 28:405-420; Bateman et al., Nucl. Acids Res. 27(1):260-262 (January 1, 1999); and Sonnhammer et al., Proteins 28(3):405-20 (July 1997).

25 From this database, consensus sequences of protein motifs and domains can be aligned with the template polypeptide sequence to determine the conserved region.

In addition, conserved regions can be determined by aligning sequences of the same or related genes in closely  
30 related plant species. Closely related plants species



preferably are from the same family. Alternatively, plant species that are both monocots or both dicots are preferred.

Sequences from two different plant species are adequate. For example, sequences from Canola and  
5 Arabidopsis can be used to identify the conserved region. Such related polypeptides from different plant species need not exhibit an extremely high sequence identity to aid in determining conserved regions.

Even polypeptides that exhibit about 35% sequence  
10 identity can be useful to identify a conserved region. More typically, such conserved regions of related proteins exhibit at least 50% sequence identity; even more typically at least about 60%; even more typically, at least 70% sequence identity, more typically at least 80%, even more  
15 typically about 90% sequence identity.

Typically, the conserved region of the target and template polypeptides or polynucleotides exhibit at least 70% sequence identity; more preferably, at least 80% sequence identity; even more preferably, at least 90%  
20 sequence identity; most preferably at least 92, 94, 96, 98, or 99% sequence identity. The sequence identity can be either at the amino acid or nucleotide level.

Sequence identity can be determined by optimal alignment of sequences to compare by the local homology  
25 algorithm of Smith and Waterman, *Add. APL. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson and Lipman, *Proc. Natl. Acad. Sci. (USA)* 85: 2444 (1988), by computerized implementations  
30 of these algorithms (GAP, BESTFIT, BLAST, PASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

Given that two sequences have been identified for comparison, GAP and BESTFIT are preferably employed to determine their optimal alignment. Typically, the default values of 5.00 for gap weight and 0.30 for gap weight length are used.

5 "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (e.g., gaps or overhangs) as compared to the reference  
10 sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched  
15 positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

Alternatively, the polynucleotides of a conserved  
20 region of closely related species will hybridize under stringent conditions wherein one of the polynucleotides is a probe to determine the conserved region. "Stringency" is a function of probe length, probe composition (G + C content), and hybridization or wash conditions of salt concentration,  
25 organic solvent concentration, and temperature. Stringency is typically compared by the parameter " $T_m$ ", which is the temperature at which 50% of the complementary The relationship of hybridization conditions to  $T_m$  (in °C) is expressed in the mathematical equation

30 
$$T_m = 81.5 - 16.6(\log_{10}[\text{Na}^+]) + 0.41(\%G+C) - (600/N) \quad (1)$$

where N is the length of the probe. This equation works well for probes 14 to 70 nucleotides in length that are identical to the target sequence. For probes of 50 nucleotides to greater than 500 nucleotides, and conditions that include an organic solvent (formamide) an alternative formulation for  $T_m$  of DNA-DNA hybrids is useful.

$$T_m = 81.5 + 16.6 \log \{ [Na^+] / (1 + 0.7[Na^+]) \} + 0.41(\%G+C) - 500/L - 0.63(\%formamide) \quad (2)$$

where L is the length of the probe in the hybrid. (P. Tijessen, "Hybridization with Nucleic Acid Probes" in Laboratory Techniques in Biochemistry and Molecular Biology, P.C. van der Vliet, ed., c. 1993 by Elsevier, Amsterdam.) With respect to equation (2),  $T_m$  is affected by the nature of the hybrid; for DNA-RNA hybrids  $T_m$  is 10-15°C higher than calculated, for RNA-RNA hybrids  $T_m$  is 20-25°C higher. Most importantly for use of hybridization to identify DNA including genes corresponding to a template sequence,  $T_m$  decreases about 1 °C for each 1% decrease in homology when a long probe is used (Bonner et al., *J. Mol. Biol.* 81:123 (1973)).

Equation (2) is derived under assumptions of equilibrium and therefore, hybridizations according to the present invention are most preferably performed under conditions of probe excess and for sufficient time to achieve equilibrium. The time required to reach equilibrium can be shortened by inclusion of a "hybridization accelerator" such as dextran sulfate or another high volume polymer in the hybridization buffer.

When the practitioner wishes to examine the result of membrane hybridizations under a variety of stringencies, an efficient way to do so is to perform the hybridization under a low stringency condition, then to wash the hybridization

membrane under increasingly stringent conditions. With respect to wash steps preferred stringencies lie within the ranges stated above; high stringency is 5-8°C below  $T_m$ .

B. Generating an Oligonucleotide

5 Once a conserved region is identified, an oligonucleotide can be generated to isolate and/or identify a target sequence. This oligonucleotide is usually not degenerate. Preferably, the oligonucleotide comprises a sequence wherein it or its reverse complement encodes a  
10 portion of the conserved region.

The portion is at least 3 amino acids in length, more typically, 4 amino acids in length; more typically at least 6 amino acids, even more typically at least 10 amino acids. Usually, the portion is at least than 40 amino acids; more  
15 usually, at least 30 amino acids; even more usually, usually at least 20 amino acids in length. A preferred range is from 3 to 18 amino acids in length.

The choice of which portion of the conserved region to use is based on convenience. Preferably, the portion of the  
20 conserved region is chosen to minimize the number of amino acids that are encoded by four or more codons. For example, the number of alanines, arginines, glycines, leucines, prolines, serines, threonines, and valines is minimized.

The sequence of the oligonucleotide is designed using  
25 the following criteria:

(1) Amino acid sequence of the conserved region of a template polypeptide;

(2) Preferred codon usage in the class, family, genera, or species of target plant species; and

30 (3) Polynucleotide sequence of the template polypeptide.

Typically, the oligonucleotide comprises at least one codon wherein the first and second position of the codon is the same as the corresponding position in the template polynucleotide and the third position is the same as the  
5 third position of the most preferred codon.

This preferred codon can be the most preferred of the plant class from which the target plant species belongs. For example, if the target plant species belongs to the dicot class, the preferred codon can be the one that is  
10 preferred by all dicots. Alternatively, the preferred codon can be one preferred in the family, genera, or species that the target plant species belongs. (The terms class, family, genera, and species is used in accordance with the accepted classification system of all organisms.)

15 One example is illustrated below:

Conserved Region (AA):           ...Aaa<sub>1</sub> - Aaa<sub>2</sub> - Aaa<sub>3</sub>...

Template Polynucleotide

encoding conserved region: (N<sub>1</sub>N<sub>2</sub>N<sub>3</sub>) - (N<sub>4</sub>N<sub>5</sub>N<sub>6</sub>) - (N<sub>7</sub>N<sub>8</sub>N<sub>9</sub>)

Preferred Codons for

20 conserved regions in

target plant species: (X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>) (X<sub>4</sub>X<sub>5</sub>X<sub>6</sub>) (X<sub>7</sub>X<sub>8</sub>X<sub>9</sub>)

Oligonucleotide: (N<sub>1</sub>N<sub>2</sub>X<sub>3</sub>) - (N<sub>4</sub>N<sub>5</sub>X<sub>6</sub>) - (N<sub>7</sub>N<sub>8</sub>X<sub>9</sub>)...

The third position of the second most preferred codon is utilized if the first two positions of the template  
25 polynucleotide do not match the most preferred codon, but the template polynucleotide matches the first two positions of the second most preferred codon.

Further, the oligonucleotide sequence is chosen to avoid homopolymers of more than four nucleotides.  
30 Preferably, a portion of the conserved region is chosen to

prevent such homopolymers from occurring in the oligonucleotide. Homopolymers can be included in the oligonucleotide if such a stretch is found in the template sequence and is preferred by the target plant species codon  
5 usage.

A higher percentage of guanosines and cytosines are preferred in the oligonucleotide sequence when a monocot target polynucleotide is to be isolated or identified using a template polynucleotide from a dicot plant species. Thus,  
10 for example, a guanosine or cytosine is preferred at the third position of the codons in the oligonucleotide when isolating and/or identifying a target sequence from a monocot using an Arabidopsis sequence as a template polynucleotide.

In contrast, higher percentage of adenines and thymidines are preferred in the oligonucleotide sequence when a dicot target polynucleotide is to be isolated or identified using a template polynucleotide from a monocot plant species. Thus, for example, an adenosine or thymidine  
15 may be preferred at the third position of the codons in the oligonucleotide when isolating and/or identifying a target sequence from a dicot, such as Arabidopsis, using a monocot sequences from corn as a template polynucleotide.

Oligonucleotides of the invention are at least 12, 16,  
25 18, 20, 25 30, 35, 40, 45 or even at least 50 nucleotides in length.

The sequence and length are chosen to generate an oligonucleotide that is capable of forming a detectable duplex with target nucleotides. The oligonucleotide can  
30 include additional nucleotides, for example inosine, that bind to sequences in the template that flank the portion of the polynucleotide encoding the conserved region to

stabilize the formed duplex. Additional non-plant polynucleotide sequences may be helpful as a label to detect the formed duplex as a primary site for PCR or to insert a restriction site for later cloning of the isolated plant sequences.

More than one oligonucleotide can be generated from the conserved region to be used in the identification and isolation procedures.

C. Isolating and/or Identifying Target Polynucleotide Sequences

The target polynucleotide sequence is isolated by contacting the oligonucleotide of the invention with a composition that comprises the target polynucleotide under conditions that permit hybridization and formation of a duplex. The duplex is then detected and the target polynucleotide can be isolated.

Exemplary procedures for identifying and/or isolating target polynucleotides that can be used include polymerase chain reaction (PCR), Southern hybridization, and polynucleotide capture.

Isolation and/or identification of a target polynucleotide can be performed using any number of oligonucleotides constructed using the instant invention.

For example, a single probe can be used in colony hybridization assays to identify from of library of clones the particular clone or clones that contain the desired target sequence. Such techniques are known, for example, for bacterial, yeast, and viral clones. Further, a single probe can also be used to generate the target polynucleotides from a starting material comprising a plurality of polynucleotides, for example in a nick

translation or cDNA synthesis or random priming or end labeling.

Single probes can be used in gel isolation techniques, such as Southern or Northern hybridization for identifying polynucleotides that correspond to the target polynucleotide to be isolated. For example, inserts of a cDNA library comprising the target polynucleotide are separated by length and are bound to a solid support so as to preserve the separation. Next, the oligonucleotide can be labeled and used to identify the fragments that hybridize to the oligonucleotide. Hybridization and wash stringency can be varied as defined above, but preferably stringent conditions are used.

Alternatively, a single oligonucleotide can be bound to a solid support to isolate the desired target polynucleotide. The solid support can be exposed to a plurality of polynucleotides. The solid support can capture those polynucleotides that hybridize to the oligonucleotide, and the unwanted polynucleotides can be washed away. The target polynucleotide can be released from the solid support and further characterized or inserted into a vector.

Other methods for capturing target polynucleotides to a solid support using an oligonucleotide are described in Li et al., U.S. Pat. No. 5,500,356; and Laffler et al., U.S. Pat. No. 5,858,652.

Oligonucleotides of the invention can be used as primers in PCR to amplify the desired target polynucleotide sequences from a plurality of polynucleotides, such as a sample of mRNA from a tissue or a cDNA library. The reaction is run using the oligonucleotides as primers and mRNA (or cDNA) or genomic DNA from the target species as a substrate. The PCR product can be inserted directly into a



vector for further processing. Alternatively, gel electrophoresis or other separations can be performed on the PCR product and the target polynucleotide can be identified by Southern hybridization techniques for further  
5 characterization or final isolation.

Amplification methods using a single oligonucleotide based on the instant invention specific for the target polynucleotide can be used for isolation and/or identification. Such a technique is single-primer PCR  
10 (SPPCR). A description of the method is described in Sreaton et al., Nucl. Acids Res. 21: 2263-2264 (1993).

Other methods of isolating target polynucleotides with a single gene specific primer are described in Frohman et al., Proc Natl Acad Sci U S A 85(23):8998-9002 (Dec. 1988)  
15 and Uematsu et al., Immunogenetics 34(3):174-8 (1991).

Also, non-specific primers comprising, for example, poly-A, poly-T, or cap sequences, can be used in conjunction with a specific oligonucleotide of the invention.

PCR amplification methods can be performed using either  
20 one or two specific oligonucleotides generated from the conserved region of the template polypeptide. Preferably, the primers generate a product that is longer than the total length of the primers. Typically, using two primers, the portions of the conserved regions that are encoded by the  
25 oligonucleotides or their reverse complements are separated by at least about 5 amino acids, more typically by at least about 30 amino acids, more typically by at least about 50 or 100 amino acids. In another acceptable arrangement, the oligonucleotides (or their reverse complements) each  
30 represent a portion of two different conserved regions of a single polypeptide. Then the polynucleotide between the

conserved regions, perhaps inclusive of one, or both of them, is amplified.

Nested primers can be used to PCR amplify the target polynucleotide sequences.

5 Compositions and methods for reverse transcriptase-polymerase chain reaction (RT-PCR) is another means of isolating and/or identifying target polynucleotides utilizing oligonucleotide primers of the invention. See, for example Lee et al, WO9844161A1 by Applicant Life  
10 Technologies.

Other amplification techniques, such as rapid amplification of cDNA ends can be used to isolate full length genes. One such procedure is described in Fehr et al., Brain Res Brain Res Protoc 3(3):242-51 (Jan. 1999).

15 D. Identifying Target Polynucleotides

The oligonucleotides of the invention can be utilized to identify the sequence of the target polynucleotides. For example, the oligonucleotides can be used in a modified PCR procedure to obtain the sequence of the target  
20 polynucleotide. See, for example, Mitchell et al., U.S. Pat. No. 5,817,797; Uhlen, U.S. Pat. No. 5,405,746; Ruano, U.S. Pat. No. 5,427,911; Leushner et al, U.S. Pat. No. 5,789,168; and

The isolated target polynucleotide can be used in any  
25 sequencing procedure, such as the known dideoxy termination method and its modifications, to identify the specific sequences.

E. Further Isolation of Target Polynucleotides

When the sequence of the target polynucleotides is  
30 identified, primers can be constructed using sequence from

the very termini of the target polynucleotides to "primer walk" and obtain the remaining sequences of the gene of which the target polynucleotides are a portion. See, for example, Screaton et al., Nucl. Acids Res. 21: 2263-2264 (1993).

The target polynucleotide can also be used to identify clones or colonies in a library that comprise sequences from the same gene as the target polynucleotide.

#### **PLANT FAMILIES**

Any plant from the plant kingdom can be used as a source of target or template polynucleotides. Without limitation, any of the plants from the monocot class, Liliopsida or from the dicot class, Magnoliopsida are of interest. Any families from these classes that can be used in the instant invention, including without limitation:

Liliaceae, Orchidaceae, Poaceae, Iridaceae, Arecaceae, Bromeliaceae, Cyperaceae, Juncaceae, Musaceae, Ameryllidaceae, Ranunculaceae, Arecaceae; Musaceae; Brassicaceae; Rosaceae; Fabaceae; Magnoliaceae; Apiaceae; Solanaceae; Lamaiaceae; Asteraceae; Salicaceae; Cucurbitaceae; Malvaceae; and Graminaceae.

Of particular interest as plants species from the following genera, without limitation, *Anacardium*, *Arachis*, *Asparagus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Manihot*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panieum*, *Pannesetum*, *Persea*, *Phaseolus*, *Pistachia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*,

*Ricinus, Secale, Senecio, Sinapis, Solanum, Sorghum, Theobromus, Trigonella, Triticum, Vicia, Vitis, Vigna, and Zea.*

**EXAMPLES**

5           The invention is illustrated by the following Examples. The invention is not limited by the Examples; the scope of the invention is defined only by the claims following.

**Example 1: General Materials and Methods**

PLANT DNAs

10           Plant DNAs were isolated according to Jofuku and Goldberg (1988); "Analysis of plant gene structure", pp. 37-66 in Plant Molecular Biology: A Practical Approach, C.H. Shaw, ed. (Oxford:IRL Press).

OLIGONUCLEOTIDES

15           Oligonucleotide primer pairs were selected from template Arabidopsis gene sequences using default parameters and the PrimerSelect 3.11 software program (Lasergene sequence analysis suite, DNASTAR, Inc., Madison, WI). Selected primer pairs were then used to generate PCR products utilizing genomic DNA from  
20 *Brassica napus* as a target plant species and polynucleotides. PCR products were either sequenced directly or cloned into *E. coli* using the TOPO<sup>TM</sup> TA vector cloning system according to manufacturer's guidelines (Invitrogen, Carlsbad, CA). Nucleotide sequences of PCR products and/or cloned inserts were determined  
25 using an ABI PRISM® 377 DNA Analyzer as specified by the manufacturer (PE Applied Biosystems, Foster City, CA) and compared to the template Arabidopsis gene sequence using default parameters and the SeqMan 3.61 software program (Lasergene sequence analysis suite, DNASTAR, Inc., Madison, WI). *Brassica napus* gene regions

of greater than or equal to 17 nucleotides in length and 70% sequence identity relative to the Arabidopsis gene were selected and the nucleotide sequences translated into the corresponding amino acid sequences using standard genetic codes. Using the deduced amino acid sequences, the corresponding sequences of triplet codons of the Arabidopsis gene region, class-, family-, genera- and/or species-specific codon usage tables, oligonucleotide primer pairs were designed for use in identifying similar gene regions that would encode identical peptides in various unrelated plant genera. In all cases, the DNA sequence of a primer or its reverse complement would be identical to the sequence of triplet codons of the Arabidopsis gene sequence at nucleotide positions 1 and 2. In some cases the nucleotide at position 3 of a triplet codon would be identical to the Arabidopsis codon if that codon is preferentially used in a given plant genera and/or species as determined by published codon usage tables. In other cases, position 3 would be selected (e.g., A, G, C, T) using genera- and/or species-specific codon usage tables such that the designated nucleotide together with nucleotides in positions 1 and 2 will form a triplet codon that will encode an amino acid that is identical to that encoded by the Arabidopsis triplet codon. In some of these cases, where there is an equal probability of using one codon or another that encodes the same amino acid but differs only at position 3, then the selection of an A, G, C, or T residue will not generate a string of homopolynucleotides more than four nucleotides.

PCR

A typical PCR reaction consisted of 1-5 µg of template plant DNA, 10 pmol of each primer of a selected primer pair, and 1.25 U of Taq DNA polymerase in standard 1X PCR reaction buffer as specified by the manufacturer (Promega, Madison, WI). PCR

reaction conditions consisted of one (1) initial cycle of denaturation at 94°C for 7 min, thirty-five (35) cycles of denaturation at 94°C for 1 min., primer-template annealing at 58°C for 30 sec., synthesis at 68°C for 4 min., and one (1) cycle of prolonged synthesis at 68°C for 7 min.

A typical single primer PCR (SPPCR) reaction consists of 1-5 µg of template plant DNA, 10 pmol of a selected primer, and 1.25 U of Taq DNA polymerase in standard 1X PCR reaction buffer as specified by the manufacturer (Promega, Madison, WI). PCR reaction conditions consisted of twenty (20) cycles of denaturation at 94°C for 30 sec., primer-template annealing at 55°C for 30 sec., synthesis at 72°C for 1 min., 30 sec., two cycles (2) of denaturation at 94°C for 30 sec., primer-template annealing at 30°C for 15 sec., 35°C for 15 sec., 40°C for 15 sec., 45°C for 15 sec., 50°C for 15 sec., 55°C for 15 sec., 60°C for 15 sec., 65°C for 15 sec., and synthesis at 72°C for 1 min., 30 sec., thirty (30) cycles of denaturation at 94°C for 30 sec., primer-template annealing at 55°C for 30 sec., synthesis at 72°C for 1 min., 30 sec., followed by one (1) cycle of prolonged synthesis at 72°C for 7 min.

#### IDENTIFICATION OF RELATED GENE SEQUENCES

Selected primers and/or primer pairs were used in PCR or SPPCR reactions using genomic DNAs isolated from selected plant genera to generate PCR products. Alternatively, primers and/or primer pairs could be used in RT-PCR reactions using RNA isolated from selected plant genera to generate PCR products using standard published procedures. PCR products were analyzed by agarose gel electrophoresis according to standard procedures. Specific products were extracted from agarose gels and either sequenced directly using the selected primer(s) as sequencing primers or first cloned into *E. coli* using the TOPO™ TA vector cloning

system according to manufacturer's guidelines (Invitrogen, Carlsbad, CA). Cloned inserts were sequenced using an ABI PRISM™ 377 DNA Analyzer as specified by the manufacturer (PE Applied Biosystems, Foster City, CA). The DNA sequences obtained were then analyzed using the MapDraw 3.15 software program (Lasergene sequence analysis suite, DNASTAR, Inc., Madison, WI). Both nucleotide and deduced amino acid sequences were then compared to the template *Arabidopsis* and *Brassica napus* gene and amino acid sequences using default parameters and the MegAlign 3.18 software program (Lasergene sequence analysis suite, DNASTAR, Inc., Madison, WI) to verify gene identity.

Alternatively, selected primers and/or PCR products could be used directly as gene probes to screen plant genomic or cDNA libraries for putative related genes in various genera and/or species. Cloned inserts identified in this way would be sequenced and the nucleotide and deduced amino acid sequences analyzed as described previously.

**EXAMPLE 2: GENERATING PRIMER SEQUENCES USING METHOD AS  
DESCRIBED -- COMPUTER SIMULATION**

**(A)**

GENE: AGAMOUS  
5 FUNCTION: TRANSCRIPTION FACTOR  
DOMAIN: MADS BOX

AA SEQUENCE: G R G K I E I K R I E  
Predicted NT: GGG AGG GGC AAG AUC GAG AUC AAG CGC AUC GAG

10 Maize GGG AGa GGC AAG AUC GAG AUC AAG CGC AUC GAG 32/33  
Rice GGG AGG GGg AAG AUC GAG AUC AAG CGg AUC GAG 31/33  
Arabidopsis GGG AGA GGA AAG AUC GAA AUC AAA CGG AUC GAG (M) 28/33  
(R) 29/33

**(B)**

GENE: APETALA1  
15 FUNCTION: TRANSCRIPTION FACTOR  
DOMAIN: MADS BOX

AA SEQUENCE: R I E N K I N R O V T F  
Predicted NT: AGG AUC GAG AAC AAG AUC AAC AAG CAG GUG ACC UUC

20 Maize cGG AUC GAG AAC AAG AUC AAC cGG CAG GUg ACC UUC 33/36  
Rice AGG AUC GAG AAC AAG AUC AAC cGG CAG GUG ACg UUC 34/36  
Arabidopsis AGG AUA GAG AAC AAG AUC AAA AGA CAA GUG ACA UUC (M) 29/36  
(R) 30/36

**(C)**

GENE: APETALA2  
25 FUNCTION: TRANSCRIPTION FACTOR  
DOMAIN: AP2 DOMAIN

AA SEQUENCE: G R W E S H I W D C  
Predicted NT: GGC AGG UGG GAG UCC CAC AUC UGG GAC UGC

30 Maize GGC cGc UGG GAa UCC CAC AUC UGG GAC UGC 27/30  
Arabidopsis GGA AGA UGG GAA UCU CAU AUU UGG GAC UGU (M) 23/30



**Example 3: SPECIFICITY OF CODON ADJUSTED PRIMERS**

The following example illustrates the specificity of codon adjusted primer pairs. Primers 1 and 2 represent primers taken directly from the sequence of the template polynucleotide. Primers 1' and 2' are primers wherein the sequence has been codon adjusted for monocots according to the invention. These primers were used to identify target polynucleotides from corn and rice.

**Primer 1**

10 AA SEQUENCE  
Coding Sequence: 5' G GAC TGT GGG AAA CAA GTT TA 3'  
Primer 1 Sequence: 5' G GAC TGT GGG AAA CAA GTT TA 3'  
  
Primer 1' (Codon Adjusted Sequence): 5' G GAC TGC GGG AAG CAG GTG TA 3' 17/21  
  
15 %Sequence Identity to Primer 1: 81%

**Primer 2**

20 AA SEQUENCE  
Coding Sequence: 5' AAG TAT AGA GGT GTC ACT TTG CA 3'  
Complement 3' TTC ATA TCT CCA CAG TGA AAC GT 5'  
  
Primer 2 Sequence: 5' TG CAA AGT GAC ACC TCT ATA CTT 3'  
  
Codon Adjusted Sequence: 5' AAG TAC AGG GGC GTC ACC TTG CA 3'  
Complement 3' TTC ATG TCC CCG CAG TGG AAC GT 5'  
  
Primer 2' Sequence: 5' TG CAA GGT GAC GCC CCT GTA CTT 3' 19/23  
  
%Sequence Identity to Primer 2: 83%

PCR was performed as described in Example 1 using genomic DNA from *Arabidopsis thaliana*, *Oryza sativa* (rice) and *Zea mays* (corn) as a source for the desired target polynucleotides.

5 **RESULTS AND CONCLUSIONS:**

PCR-amplified products of the expected size were generated using primers 1 and 2 and *Arabidopsis* genomic DNA as a substrate. No products were obtained in reactions using either rice or corn genomic DNA substrate.

10 On the other hand, PCR-amplified products were generated using the codon adjusted primers 1' and 2' and corn genomic DNA as a substrate. No products were obtained in a reaction using *Arabidopsis* genomic DNA substrate. Together, these results demonstrate the general utility of  
15 designing codon adjusted primers for use in isolating/identifying gene orthologs from different plant families.

**Example 4**

The method of the invention was used to isolate AP2-like genes from *Avena sativa* (oat), *Oryza sativa* (rice), *Triticum aestivum* (wheat) and *Zea mays* (corn). Primers 1' and 2' described in Example 3 were used in PCR using the conditions of Example 1 and genomic DNA from each plant as a source of target polynucleotides. The nucleotide and corresponding amino acid sequences of PCR-amplified products are shown below.

>OAT ADC GENE 489 BP

TACCTAGGTGAGCTCAAATTTCCAGCTCCAGCTCCTCCTAATTAATTTCCATCTGTTCTGTGTACTGAAGTTATTTAATTTTCGTACGGTGG  
TTTCGACACCGCGCACTCGGCGCGAGGTTATAATTAATCAAGCTTCTAGTTTGAACCTTTCAACACATACTGCTCTCTCTCGATTGGATT  
GTACTAGCATCATGAAGTGTACTGAAACGGGTCTTGCTCAGGGCTACGATCGCGCGGCGATCAAGTTCCGGGGACTGGACGCCGACATCA  
ACTTCAATCTGAGCGACTACGAGGAGGATCTGAAGCAGGTAAGTGAATAAGATCGCTTCTCAAATGCAGCATAGATATTATCGGTGTGTG  
TGTGTCTGATGGGTGGTGGTGGCGCGCGGCACTCTTGTCTTGGCAGATGAGGAAGTGGACCAAGGAGGAGTTCGTGCACATCCTCCG  
CCGCCAGAGCACGGGGTTTCGCGAGGGGAGCTCA

>OAT ADC PROTEIN 65 aa

GGFDTAHSAARAYDRAAIKFRGLDADINFNLSDYEDLKQVTNWTKEEFVHILRRQSTGFARGSS

>RICE AP2-LIKE GENE 387 BP

CCTAGGTAATTTTCATCGAACACATCATCTTCCTCCTCTCAATCCAACGCGACATCGCCATGAACAATCTAACAAACACCTTCATCTTCTCC  
CAAACAATCACAGGTGGATTTCGACACTGCTCACGCAGCTGCAAGGTAAGAAACACATCACATCATTATCAGAACATGAGCTCTGTGTTTG  
TGAAGGAGATTGAGAGAATTGAATGATGATGGATGGATGCAGGGCGTACGACAGGGCGGCGATCAAGTTCCAGGGAGTAGAGGCTGACATC  
AACTTCAACCTGAGCGACTACGAGGAGGACATGAGGCAGATGAAGAGCTTGTCCAAGGAGGAGTTCGTGCACGTTCTCCGGCGACAGAGCA  
CCGGCTTCTCCCGCGGCAGCTCA

>RICE ADC PROTEIN 65 aa

GGFDTAHAAARAYDRAAIKFRGVEADINFNLSDYEDDMRQMKSLSKEEFVHVLRRQSTGFSRGSS

>WHEAT ADC GENE 477 BP

CTTGGGTGGGTTTGACACTGCACATGCTGCTGCAAGGTACGTACAAATTTAATTAAGCACGTACGCAGTACATAATTGTGATGTGATCATC  
ACCTGAACCACTGTACTGCAACTCTGAAGTTATGTCTCCACTCTGTTTCATTTCACCGTGCCAAATTGACCTTGGGATGTTCCGCGAGGGCG  
TACGATCGAGCGGCGATCAAGTTCCGCGCGTGCACGCCGACATAAACTTCAACCTCAGCGACTACGAGGACGACATGAAGCAGGTGATCA  
GCAAAGCCACCAACAGTGTTCCTCATCAACCAAATTTATTCAGATGCAGAGTGCATTAGTACTGTTGTTGAACTGATGAACTGAAGAAA  
TTCTGACTGTGTGTTGKTTGGTGGATGATCTGGATCAGATGAAGGGCTGTCCAAGGAGGAGTTCGTGCACGTGCTGCGGCGGACAGAGCGC  
CGGCTTCTCGCGGGCAGCTCC

>WHEAT ADC PROTEIN 65 aa

GGFDTAHAAARAYDRAAIKFRGVDADINFNLSDYEDDMQVKGLSKEEFVHVLRRQSAGFSRGSS

>MAIZE ADC GENE 489 BP

CTTAGGTGAGCAGCAATAAGCAGATCGATCTGCAGCATAAATTTCCCGTTATTAAGTCTGATCTCGATCGAATGGCCTAATTAACC  
GATTCCGGTATCTGGCCGATGGCCAATCTACGCAGGTGGATTTCGACACTGCTCATGCCGCTGCAAGGTAACGATCAATCCATCCATCCACC  
CTTGTCTAGCTACCCACCGACCGCGGATTATGGACCGTAGTTCTCGGACGGGCTTGCTGCAGGGCGTACGACCGAGCGGCGATCA  
AGTTCCGCGGCGTGCAGCCGACATAAACTTCAACCTCAGCGACTACGACGACGATATGAAGCAGGTACATACAGAGTGTGTTGTCAGCT  
AGCACCGACTGAAACATCTGCTGAACGTACACTTACGGCCTGTGCACCGATGAAGAGCTGTCCAAGGAGGAGTTCGTGCACGCCCTGCG  
GCGGCAGAGCACCGGCTTCTCCGTTGGCAGCTCC

>MAIZE ADC PROTEIN 65 aa

GGFDTAHAAARAYDRAAIKFRGVDADINFNLSDYDDMKQVKLSKEEFVHALRRQSTGFSRGSS

**EXAMPLE 5. USE OF SHORT CODON ADJUSTED PRIMERS**

*Oligonucleotides*

Codon adjusted oligonucleotides were designed as described previously. Derivatives of oligonucleotide 2' were generated as shown above and used as primers in combination with oligonucleotide 1' in PCR reactions using plant genomic DNA from Zea mays (corn), Avena sativa (oat), and Triticum aestivum (wheat) as a source of target polynucleotides.

*PCR*

A typical PCR reaction consisted of 1-5 µg of target plant DNA, 10 pmol of primer 1' and 10 pmol of a derivative of primer 2', and 1.25 U of Taq DNA polymerase in standard 1X PCR reaction buffer as specified by the manufacturer (Promega, Madison, WI). PCR reaction conditions consisted of five cycles (5) of denaturation at 94oC for 2 minutes, 94oC for 30 sec., primer-template annealing at 65oC for 15 sec., 60oC for 15 sec., 55oC for 15 sec., 50oC for 15 sec., 45oC for 15 sec., 40oC for 15 sec., and synthesis at 68oC for 1 min., 30 sec., and twenty (20) cycles of denaturation at 94oC for 30 sec., primer-template annealing at 55oC for 30 sec., synthesis at 72oC for 1 min., 30 sec., thirty (30) cycles of denaturation at 94oC for 30 sec., primer-template annealing at 50oC for 30 sec., synthesis at 68oC for 1 min., followed by one (1) cycle of prolonged synthesis at 68oC for 7 min.

25 Primer 1

AA SEQUENCE

Coding Sequence:

Primer Sequence:

D C G L Q V  
5' G GAC TGT GGG AAA CAA GTT TA 3'  
5' G GAC TGT GGG AAA CAA GTT TA 3'

Primer 1' (Codon Adjusted Sequence):

5' G GAC TGC GGG AAG CAG GTG TA 3'

25

30 Primer 2

AA SEQUENCE

Coding Sequence:

Complement

K Y R G V T L

5' AAG TAT AGA GGT GTC ACT TTG CA 3'

3' TTC ATA TCT CCA CAG TGA AAC GT 5'

Primer 2 Sequence:

5' TG CAA AGT GAC ACC TCT ATA CTT 3'

35 Codon Adjusted Sequence:

Complement

5' AAG TAC AGG GGC GTC ACC TTG CA 3'

3' TTC ATG TCC CCG CAG TGG AAC GT 5'

Primer 2' Sequence:

5' TG CAA GGT GAC GCC CCT GTA CTT 3'

RISZU2'-1 (5 CODONS)

5' G CAA GGT GAC GCC CCT GT 3'

RISZU2'-2 (5 CODONS)

5' GGT GAC GCC CCT GTA CT 3'

RISZU2'-3 (4 CODONS)

5' GT GAC GCC CCT GTA CT 3'

RISZU2'-4 (3 CODONS)

5' GT GAC GCC CCT GT 3'

RESULTS AND CONCLUSIONS:

As described in Methods, primer 2' derivatives vary in length from 15-18 bp that could encode a peptide of 4-5 amino acids in length. Figure 3 shows that PCR-amplified products were generated using primer 1' and primer 2' derivatives 1, 2, and 3 and all three genomic DNAs as a source of target polynucleotides.

These results demonstrate that the method as described can utilize conserved regions of greater than or equal to 4 amino acids in length for use in isolating/identifying gene orthologs from different plant families.

**CLAIMS**

What is claimed:

1. A method for isolating a from a target plant species a target polynucleotide encoding a target polypeptide comprising a conserved region exhibiting at least 70% sequence identity to a conserved region of template polypeptide that is encoded by a template polynucleotide from a template plant species,  
5 comprising:

(a) identifying an amino acid sequence of a conserved region in the template polypeptide;

(b) generating an oligonucleotide comprising a sequence  
10 wherein the sequence or its reverse complement comprises at least four codons that encode a portion of the amino acid sequence of (a), wherein

(i) the sequence of the first and second positions of at least three of the codons is the same as the corresponding nucleotides in nucleotides in the template polynucleotide;  
5

(ii) the nucleotide at the third position of the codons of (i) is the nucleotide of the third position of the most preferred codon of the target plant class for the desired amino acid;  
20

(c) contacting the oligonucleotide with a composition comprising the target polynucleotide under conditions that permit hybridization of the oligonucleotide to the target polynucleotide to form a duplex; and

(d) isolating the duplex.  
25

2. The method of claim 1, wherein the oligonucleotide does not contain a homopolymer of more than four guanine or cytosine residues.

3. The method of claim 1, wherein the oligonucleotide does not contain a homopolymer of more than four residues.

4. The method of claim 1, wherein the oligonucleotide of step (b) wherein the sequence or its reverse complement further comprises at least one codon wherein

(i) the sequence of the first and second position of  
5 the codon is the same as the corresponding nucleotides in the template polynucleotide;

(ii) the sequence of the third position of the codon of  
step (I) is the same as the nucleotide of the third position  
of the second most preferred codon of the target plant  
10 species for the desired amino acid; and

(iii) the oligonucleotide is not degenerate.

5. The method of claim 1, wherein the target polynucleotide is from a monocot plant and the template polynucleotide is from a dicot plant.

6. The method of claim 4, wherein the template polynucleotide is from Arabidopsis.

7. The method of claim 5, wherein the third position of each codon is either a guanosine or cytosine.

8. The method of claim 2, wherein both the template and target polynucleotides are from dicot plants.

9. The method of claim 8, wherein the template polynucleotide is from Arabidopsis.

10. The method of claim 9, wherein the third position of each codon is either an adenosine or thymidine.

11. The method of claim 1, wherein the template polynucleotide is from a monocot plant and the target polynucleotide is from a dicot plant.

12. The method of claim 1, wherein both the template and target polynucleotides are from monocot plants.

13. The method of claim 11 or 12, wherein the template polynucleotide is from corn.

14. The method of claim 12, wherein the target polynucleotide is from corn.

15. The method of claim 1, wherein step (a) comprises aligning polynucleotides of plants within a family and identifying a portion of the template polynucleotide that exhibits at least 70% sequence identity to a portion of a polynucleotide from a plant of a genus closely related to the plant from which the template polynucleotide originates.

16. The method of claim 1, wherein step (a) comprises identifying the primary sequence in a region of the template polypeptide sequence that forms a secondary structure.

17. The method of claim 16, wherein the secondary structure is a helix or a beta sheet.

18. The method of claim 1, wherein the conserved region is a motif or functional domain.



19. The method of claim 1, wherein step (a) comprises identifying the primary sequence in a region of the template polypeptide that is repeated.

20. The method of claim 1, wherein the oligonucleotide comprises from 6 to 11 codons.

21. The method of claim 1, wherein step (c) further includes contacting the composition comprising the target polynucleotide with a second oligonucleotide, wherein the second oligonucleotide is a degenerate oligonucleotide  
5 encoding a second portion of the conserved region.

22. A method of isolating a target polynucleotide encoding a conserved region in a template polypeptide encoded by a template polynucleotide comprising:

(a) identifying the amino acid sequence of the  
5 conserved region in the template polypeptide;

(b) generating an oligonucleotide comprising a sequence wherein the sequence or its reverse complement comprises at least four codons that encode a portion of the conserved region of step (a), wherein

10 (i) the sequence of the first and second positions of at least three codons is the same as the corresponding nucleotides in the template polynucleotide;

15 (ii) the nucleotide of the third position of those six codons is the same nucleotide in the third position of the most preferred codon of the target plant species for the desired amino acid;

(iii) the oligonucleotide does not comprise homopolymers of more than four nucleotides; and

20 (iv) the oligonucleotide is not degenerate;

(c) contacting the oligonucleotide with a composition comprising the target polynucleotide under conditions that permit hybridization of the oligonucleotide to the target polynucleotide to form a duplex;

25 (d) contacting the duplex of step (c) with a thermostable polymerase under conditions to elongate the oligonucleotide of step (b); and

(e) isolating the elongation product of step (d).

23. A method for identifying a target polynucleotide encoding a conserved region in a template polypeptide encoded by a template polynucleotide comprising:

(a) identifying the amino acid sequence of the  
5 conserved region in the template polypeptide;

(b) generating an oligonucleotide comprising a sequence wherein the sequence or its reverse complement comprises four codons that encode a portion of the conserved region of step (a), wherein

10 (i) the sequence of the first and second positions of at least three codons is the same as the corresponding nucleotides in the template polynucleotide;

(ii) the nucleotide of the third position of those  
15 six codons is the same nucleotide in the third position of the most preferred codon of the target plant species for the desired amino acid;

(iii) the oligonucleotide does not comprise homopolymers of more than four nucleotides; and

20 (iv) the oligonucleotide is not degenerate;

(c) contacting the oligonucleotide with a composition comprising the target polynucleotide under conditions that permit hybridization of the oligonucleotide to the target polynucleotide to form a duplex;

25 (d) contacting the duplex of step (c) with a thermostable polymerase under conditions to elongate the oligonucleotide of step (b); and

(e) determining the nucleotide sequence of the elongation product of step (d).

24. A method of isolating a target polynucleotide encoding a polypeptide of a conserved region in a template polypeptide encoded by a template polynucleotide, comprising:

5 (a) identifying the amino acid sequence of the conserved region in the template polypeptide;

(b) generating a first oligonucleotide comprising a sequence wherein the sequence or its reverse complement comprises four codons that encode a first portion of the  
10 conserved region of step (a), wherein

(i) the sequence of the first and second position of at least three codons is the same as the corresponding nucleotides in the template polynucleotide;

15 (ii) the nucleotide of the third position of the codons of step (i) is the same as the nucleotide in the third position of the most preferred codon of the target plant species for the desired amino acid;

(iii) the oligonucleotide does not comprise  
20 homopolymers of more than four nucleotides; and

(iv) the oligonucleotide is not degenerate;

(c) generating a second oligonucleotide wherein its sequence or its reverse complement comprises four codons that encode a second portion of the conserved region of step

25 (a), wherein

(i) the sequence of the first and second position of at least three codons is the same as the corresponding position in the template polynucleotide;

30 (ii) the nucleotide of the third position of those codons is the same as the nucleotide of the third position of the most preferred codon of the target plant species for the desired amino acid;

(iii) the oligonucleotide does not comprise homopolymers of more than four nucleotides; and

35 (iv) the oligonucleotide is not degenerate;

(d) contacting the first and second oligonucleotides with a composition comprising the target polynucleotide under conditions that permit hybridization of at least one of the oligonucleotides and the target polynucleotide to  
40 form a duplex;

(e) contacting the duplex of step (d) with a thermostable polymerase under conditions to elongate the at least one hybridized oligonucleotide;

45 (f) generating a strand complementary to the elongation product of step (e); and

(g) isolating the product of step (d).

25. The method of claim 24, wherein the two oligonucleotide sequences or their reverse complements encode portions of the conserved region of step (a) that are separated by at least 30 amino acids.

26. The method of claim 24, wherein the two oligonucleotide sequences or their reverse complement encode between 6 to 11 amino acids of the conserved region of step (a).

27. The method of claim 24, wherein the product of step (f) is inserted into a vector.

28. A method for identifying a target polynucleotide encoding a polypeptide of a conserved region in a template polypeptide encoded by a template polynucleotide, comprising:

5 (a) identifying the amino acid sequence of the conserved region in the template polypeptide;

(b) generating a first oligonucleotide comprising a sequence wherein the sequence or its reverse complement comprises four codons that encode a first portion of the  
10 conserved region of step (a), wherein

(i) the sequence of the first and second position of at least three codons is the same as the corresponding nucleotides in the template polynucleotide;

15 (ii) the nucleotide of the third position of the codons of step (i) is the same as the nucleotide in the third position of the most preferred codon of the target plant species for the desired amino acid;

(iii) the oligonucleotide does not comprise  
20 homopolymers of more than four nucleotides; and

(iv) the oligonucleotide is not degenerate;

(c) generating a second oligonucleotide wherein its sequence or its reverse complement comprises four codons

that encode a second portion of the conserved region of step  
25 (a), wherein

(i) the sequence of the first and second position  
of at least three codons is the same as the  
corresponding position in the template polynucleotide;

30 (ii) the nucleotide of the third position of those  
codons is the same as the nucleotide of the third  
position of the most preferred codon of the target  
plant species for the desired amino acid;

(iii) the oligonucleotide does not comprise  
homopolymers of more than four nucleotides; and

35 (iv) the oligonucleotide is not degenerate;

(d) contacting the first and second oligonucleotides  
with a composition comprising the target polynucleotide  
under conditions that permit hybridization of at least one  
of the oligonucleotides and the target polynucleotide to  
40 form a duplex;

(e) contacting the duplex of step (d) with a  
thermostable polymerase under conditions to elongate the at  
least one hybridized oligonucleotide;

45 (f) generating a strand complementary to the  
elongation product of step (e); and

(g) determining the nucleotide sequence of the product  
of step (f).

29. A method for selecting a nucleotide sequence of an  
oligonucleotide primer for a polymerase chain reaction  
comprising:

5 (a) selecting a nucleotide sequence encoding a desired  
amino acid sequence from a template organism, or the  
complement thereof;

(b) selecting for the nucleotide of the third position of each codon the preferred codon for a target organism, provided said nucleotide is guanine or cytosine;

10 (c) if the nucleotide of the third position of the preferred codon is adenine or thymine, then substituting either a guanine or cytosine, selecting guanine or cytosine to avoid introducing a poly-guanylate or polycytidylate sequence of more than four residues;

15 wherein said desired amino acid sequence is encoded by one reading frame, or a portion thereof, of the nucleotide sequence of said primer or the complement thereof.

30. A method for preparing an oligonucleotide primer for a polymerase chain reaction comprising:

(a) selecting a nucleotide sequence encoding a desired amino acid sequence from a template organism, or the  
5 complement thereof;

(b) selecting for the nucleotide of the third position of each codon the preferred codon for a target organism, provided said nucleotide is guanine or cytosine;

10 (c) if the nucleotide of the third position of the preferred codon is adenine or thymine, then substituting either a guanine or cytosine, selecting guanine or cytosine to avoid introducing a poly-guanylate or polycytidylate sequence of more than four residues; and

(d) synthesizing said oligonucleotide primer, wherein  
15 said desired amino acid sequence is encoded by one reading frame, or a portion thereof, of the nucleotide sequence of said primer or the complement thereof.

31. A method for cloning a nucleic acid comprising:

(a) selecting an upstream nucleotide sequence encoding a first desired amino acid sequence from a template organism and a downstream nucleotide sequence encoding a second  
5 desired amino acid sequence;

(b) for each of said upstream and downstream nucleotide sequences, selecting for the nucleotide of the third position of each codon the preferred codon for a target organism, provided said nucleotide is guanine or  
10 cytosine;

(c) if the nucleotide of the third position of the preferred codon is adenine or thymine, then substituting either a guanine or cytosine, selecting guanine or cytosine to avoid introducing a poly-guanylate or polycytidylate  
15 sequence of more than four residues.

32. The method of claim 31, further comprising:

(d) synthesizing an upstream oligonucleotide primer, or a portion thereof according to steps (b) and (c).

33. The method of claim 32, further comprising:

(e) performing a polymerase chain reaction using said upstream and downstream primers and a template comprising a  
5 nucleic acid sample obtained from said target organism.

34. The method of claim 33, further comprising:

(f) using the product of said polymerase chain reaction of step (e) as a probe to screen a library prepared from nucleic acids obtained from said target organism.

35. The method of claim 33, further comprising:



(f') inserting the product of the polymerase chain reaction of step (e) into a vector.

36. The method of any one of claims 30-35, wherein said template organism is a dicot and said target organism is a monocot or wherein said template organism is a monocot and said target organism is a dicot.

37. A method for isolating a target polynucleotide encoding a target polypeptide comprising a conserved region of a template polypeptide that is encoded by a template  
5 polynucleotide, comprising:

(a) identifying an amino acid sequence of a conserved region in the template polypeptide;

(b) generating an oligonucleotide comprising a sequence wherein the sequence or its reverse complement  
10 comprises at least four codons that encode a portion of the amino acid sequence of (a), wherein

(i) the sequence of the first and second positions of at least three of the codons is the same as the corresponding nucleotides in  
15 nucleotides in the template polynucleotide;

(ii) the nucleotide at the third position of the codons of (i) is the nucleotide of the third position of the most preferred codon of the target plant species for the desired amino acid;

(c) contacting the oligonucleotide with a composition  
20 comprising the target polynucleotide under conditions that permit hybridization of the oligonucleotide to the target polynucleotide to form a duplex; and

(d) generating a single strand polynucleotide.

38. A method for isolating a from a target plant species a target polynucleotide encoding a target polypeptide comprising a conserved region exhibiting at least 70% sequence identity to a conserved region of template polypeptide that is encoded by a template polynucleotide from a template plant species, comprising:

(a) identifying an amino acid sequence of a conserved region in the template polypeptide;

(b) generating an oligonucleotide comprising a sequence wherein the sequence or its reverse complement comprises at least four codons that encode a portion of the amino acid sequence of (a), wherein

(i) the sequence of the first and second positions of at least three of the codons is the same as the corresponding nucleotides in nucleotides in the template polynucleotide;

(ii) the nucleotide at the third position of the codons of (i) is the nucleotide of the third position of the most preferred codon of the plant family of target plant species for the desired amino acid;

(c) contacting the oligonucleotide with a composition comprising the target polynucleotide under conditions that permit hybridization of the oligonucleotide to the target polynucleotide to form a duplex; and

(d) isolating the duplex.

39. A method for isolating a from a target plant species a target polynucleotide encoding a target polypeptide comprising a conserved region exhibiting at

5 least 70% sequence identity to a conserved region of  
template polypeptide that is encoded by a template  
polynucleotide from a template plant species, comprising:

(a) identifying an amino acid sequence of a conserved  
region in the template polypeptide;

10 (b) generating an oligonucleotide comprising a  
sequence wherein the sequence or its reverse complement  
comprises at least four codons that encode a portion of the  
amino acid sequence of (a), wherein

15 (i) the sequence of the first and second  
positions of at least three of the codons is the  
same as the corresponding nucleotides in  
nucleotides in the template polynucleotide;

20 (ii) the nucleotide at the third position of  
the codons of (i) is the nucleotide of the third  
position of the most preferred codon of the plant  
genera for the target plant species for the  
desired amino acid;

25 (c) contacting the oligonucleotide with a composition  
comprising the target polynucleotide under conditions that  
permit hybridization of the oligonucleotide to the target  
polynucleotide to form a duplex; and

(d) isolating the duplex.

40. A method for isolating a from a target plant  
species a target polynucleotide encoding a target  
polypeptide comprising a conserved region exhibiting at  
5 least 70% sequence identity to a conserved region of  
template polypeptide that is encoded by a template  
polynucleotide from a template plant species, comprising:

(a) identifying an amino acid sequence of a conserved region in the template polypeptide;

10 (b) generating an oligonucleotide comprising a sequence wherein the sequence or its reverse complement comprises at least four codons that encode a portion of the amino acid sequence of (a), wherein

15 (i) the sequence of the first and second positions of at least three of the codons is the same as the corresponding nucleotides in nucleotides in the template polynucleotide;

20 (ii) the nucleotide at the third position of the codons of (i) is the nucleotide of the third position of the most preferred codon of the plant species of the target plant species for the desired amino acid;

25 (c) contacting the oligonucleotide with a composition comprising the target polynucleotide under conditions that permit hybridization of the oligonucleotide to the target polynucleotide to form a duplex; and

(d) isolating the duplex.

**ABSTRACT**

The invention provides an improved method for isolating or characterizing genes in a second or target plant species that have substantial sequence identity to at least a portion of a gene in a first or template plant species.

POWER OF ATTORNEY

CERES, INC.  
3007 Malibu Canyon Road  
Malibu, CA 90265

I, Richard Hamilton, Chief Financial Officer of CERES, INC. of 3007 Malibu Canyon Road, Malibu, California 90265, grant Power of Attorney and authority to empower the following attorneys to act on behalf of CERES, INC. for executing Verified Statements (Declarations) Claiming Small Entity Status to be submitted to the U.S. Patent and Trademark Office in connection with the filing of provisional or regular patent applications on behalf of CERES, INC.

Raymond C. Stewart (Reg. No. 21,066)  
Joseph A. Kolasch (Reg. No. 22,463)  
Leonard R. Svensson (Reg. No. 30,330)  
Gerald M. Murphy, Jr. (Reg. No. 28,977)  
Mark J. Nuell (Reg. No. 36,623)

This Power of Attorney is to remain in full force and effect until terminated by an official of CERES, INC.

By   
Richard Hamilton

Date 9/24/98